

cd2sbgnml: bidirectional conversion between CellDesigner and SBGN formats

Irina Balaur^{1,*}, Ludovic Roy^{2,3,+}, Alexander Mazein^{1,4,5*}, S. Gökberk Karaca⁶, Ugur Dogrusoz⁶, Emmanuel Barillot^{2,3,7}, Andrei Zinovyev^{2,3,7}

¹ European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Université de Lyon, 50 Avenue Tony Garnier, 69007 Lyon, France

² Institut National de la Santé et de la Recherche Médicale, U900, Paris, France

³ MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France

⁴ Institute of Cell Biophysics, Russian Academy of Sciences, Pushchino, 142290, Russia

⁵ Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, L-4367 Belvaux, Luxembourg

⁶ Computer Engineering Department, Bilkent University, Ankara 06800, Turkey

⁷ Institut Curie, PSL Research University, Paris, France

*Corresponding authors: irinaa.balaur@gmail.com , a.mazein@gmail.com + Equal contribution

Abstract

Motivation: CellDesigner is a well-established biological map editor used in many large-scale scientific efforts (Funahashi *et al.*, 2007). However, the interoperability between the Systems Biology Graphical Notation (SBGN) Markup Language (SBGN-ML) and the CellDesigner file format remains a challenge due to the proprietary extensions used in CellDesigner files.

Results: We introduce a library named cd2sbgnml and an associated web service for bidirectional conversion between CellDesigner's proprietary Systems Biology Markup Language (SBML) extension and SBGN-ML formats. We discuss the functionality of the cd2sbgnml converter, which was successfully used for the translation of comprehensive large-scale diagrams such as the RECON Human Metabolic network (Noronha *et al.*, 2019) and the complete Atlas of Cancer Signalling Network (Kuperstein *et al.*, 2015), from the CellDesigner file format into SBGN-ML.

Availability: The cd2sbgnml conversion library and the web service were developed in Java, and distributed under the GNU Lesser General Public License v3.0. The sources along with a set of examples are available on GitHub (<https://github.com/sbgn/cd2sbgnml> and <https://github.com/sbgn/cd2sbgnml-webservice>, respectively).

Contact: cd2sbgnml-dev@googlegroups.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Introduction

Systems Biology standard formats such as Systems Biology Markup Language (SBML) (Hucka *et al.*, 2003), System Biology Graphical Notation (SBGN) (Le Novère *et al.*, 2009) and Biological Pathways Exchange Language (BioPAX) (Demir *et al.*, 2010) have been developed to allow accurate computational description of biological systems and to assure complete interoperability for biological resources and full portability for biomodels. The SBGN standard focuses on a graphical representation. It includes three complementary languages

for diagram representation, namely i) the Process Description (PD) to describe biochemical interactions, ii) the Activity Flow (AF) to represent information flow among biochemical entities and iii) the Entity Relationship (ER) to illustrate relationships in which given entities participate in biological networks.

CellDesigner is a well-established Systems Biology Workbench that facilitates biological network management including diagram editing and mathematical exploration of biochemical interactions (Funahashi *et al.*, 2007). Maps developed using CellDesigner are included in large-scale scientific efforts such as the PANTHER Pathway (Mi and Thomas, 2009), the BioModels (Le Novère *et al.*, 2006), the Atlas of Cancer Signaling Network (Kuperstein *et al.*, 2015) and the Virtual Metabolic Human (Noronha *et al.*, 2019) databases. However, the interoperability between the SBGN standard format and the CellDesigner file format (SBML extended with layout information) remains challenging given that CellDesigner uses a proprietary format. For example, in addition to generic protein (corresponding to macromolecule in the SBGN PD), CellDesigner also has receptor, ion channel, and truncated protein glyphs. In addition to state transition (corresponding to the generic process glyph in SBGN), CellDesigner has specific types of processes including translation, transcription, transport, and truncation. Therefore, there is a need for translation between the CellDesigner format and SBGN standards. Here, we present the cd2sbgnml library for two-way conversion of the CellDesigner file format and SBGN-ML. We also introduce a web service running this conversion library for use by systems biology software tools.

Methods

The cd2sbgnml converter is a standalone Java-based software developed for translation between the CellDesigner (version 4.4) file format, which is an extension of SBML with layout information, and the PD and AF languages of the SBGN standard format. The cd2sbgnml tool uses i) the libSBGN library (van Iersel *et al.*, 2012) dedicated to management (reading, writing) of the SBGN files and ii) the JAXB library and a manually-curated XSD file (created specifically for CellDesigner 4.4) for handling CellDesigner files. The cd2sbgnml dual converter offers both a command line utility (scripts) and a graphical user interface to facilitate access to its functionality. A log file is also created to accommodate eventual messages of exceptions and warnings generated during the translation process.

Information on notes and annotations is preserved from CellDesigner to SBGN for the model, species, reactions and compartments. In addition, note elements are taken from the celldesigner:species, celldesigner:protein, celldesigner:gene, celldesigner:RNA and celldesigner:AntisenseRNA elements. For the SBGN to CellDesigner translation, information on notes and annotations is taken from maps and glyphs only but not arcs. Specifically, process, map and glyph annotations are used for <reaction>, <model> and <species> elements, respectively. Given that CellDesigner does not allow annotations for the individual components of complexes (called included species), information on RDF annotations of such elements is ignored during the translation. Finally, information on the style of nodes and edges (including color and annotation) is also conserved in both directions.

Managing SBGN to CellDesigner translation

SBGN process glyph representation during the translation process. A biological process is represented in SBGN by the process glyph that includes incoming and outgoing links (arcs)

for consumption and production, respectively, and eventually, regulatory links (arcs) such as catalysis and inhibition. The connection between the process glyph and the corresponding incoming/outgoing arcs is made via SBGN ports; specifically, each process glyph has two ports, one for incoming and another for outgoing arcs. Given the fact that CellDesigner does not provide a specific representation for the SBGN ports, the cd2sbgnml converter introduced two graphical points associated with each process shape for an accurate illustration of the SBGN process glyphs (as shown in Supplementary file 1). The SBGN ports corresponding to the SBGN logical operators are treated similarly by the converter.

SBGN submap and perturbing agent glyphs. Given that the SBGN submap and perturbing glyphs are not addressed in CellDesigner notation, both glyphs are converted to phenotypes. The terminals inside the submaps are not translated, but the arcs pointing to each terminal are set to point to the replacing phenotype glyph. Those equivalence arcs are translated to instances of POSITIVE_INFLUENCE.

Managing CellDesigner to SBGN-ML translation

CellDesigner-specific protein-related representations such as the active/inactive states (active entities have a dashed line on the outside of their shape), the hypothetical entities (entities with a dashed border) and the binding region box are not translated since they have no equivalent in SBGN. Also, the number of a multimer unit greater than 2 is illustrated graphically in CellDesigner, but only numerically in SBGN (e.g. N:unit count).

Ports. Given that the port length is fixed, automatic port generation can cause changes for the arc orientation, introducing therefore some aesthetics, especially in compact networks, depending on how the SBGN is rendered. A more readable network can be obtained if enough space is left between every node.

The CellDesigner unknown logic gate have no correspondent in SBGN; in this case, the cd2sbgnml converter links directly the concerned input entities to the output process.

Direct links to other links. In CellDesigner, arcs can point to other arcs. However, this representation is valid in SBGN Entity Relationship only, which is not currently supported by the cd2sbgnml converter. Thus, such links are discarded by the translator.

Representation of CellDesigner “mixed” diagrams. While the reduced notation of CellDesigner is similar to the AF SBGN language, CellDesigner allows development of “mixed” diagrams, where the main notation and reduced notation are combined. However, given that the SBGN standard has separate means to represent PD and AF projections, losing reactions of this or that type is unavoidable when converting such mixed diagrams.

cd2sbgnml as a web service

The cd2sbgnml conversion library can be easily turned into a webservice through another library named cd2sbgnml-service. An example usage of this service can be found in Newt (release 1.1.0, available since June 21, 2018) under the File import and export menus. Newt (<http://web.newteditor.org>) is a free, web based, open source viewer and editor for pathways in the SBGN standard format.

Results

The cd2sbgnml converter was used to translate from CellDesigner’s proprietary SBML extension into SBGN a set of biological maps including those of the RECON Human Metabolic network (Noronha *et al.*, 2019) and of the Atlas of Cancer Signalling Network

(ACSN) (Kuperstein *et al.*, 2015). These maps are among the most comprehensive biological diagrams to date, available to the Systems Biology scientific community. For example, the Recon3D map contains 17030 species (4140 unique metabolites and 12890 proteins) and 13543 metabolic reactions (Noronha *et al.*, 2019) and the complete ACSN 2.0 map is composed of 13 interconnected signaling maps, containing around 9692 species (2997 proteins, 740 RNAs, 130 antisense RNAs, 808 genes, 665 simple molecules, 775 small molecules) and 8137 reactions (<https://acsn.curie.fr/>). Conversion from the CellDesigner file format to SBGN, using a MacBook Pro (3 GHz Intel Core i7, 16 GB), took ~13 sec. for Recon3D and ~18 sec. for the complete ACSN map. Details on map composition (number of species and reactions/processes involved) and conversion time for these map collections are given in Table 1 of Supplementary file 2. Figure 1 includes a simple illustration on the receptor representation in A) CellDesigner and in SBGN within B) the Newt editor and C) VANTED (Rohn *et al.*, 2012), (based on the ACSN Dendritic Cell map: <https://acsn.curie.fr/>). Specifically, receptor information is represented graphically as a shape with 6 sides in CellDesigner and is given by the unit of information glyph attached to the corresponding protein in SBGN.

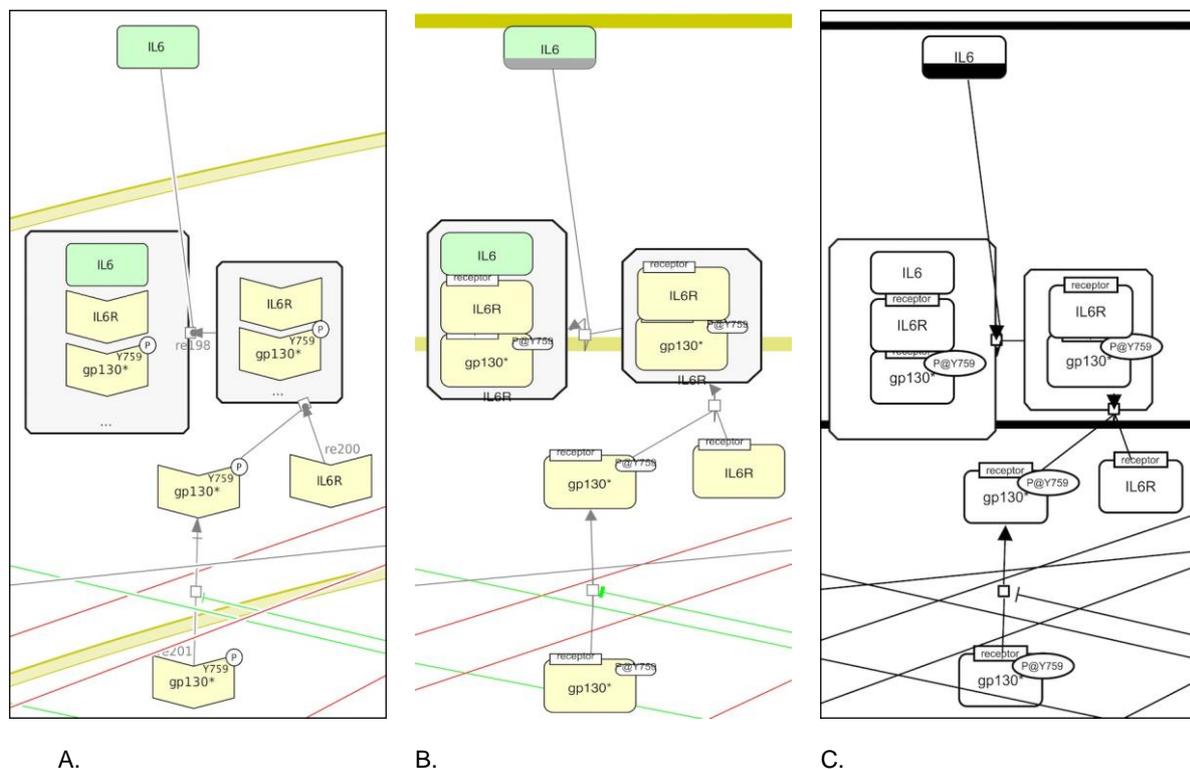


Figure 1. Simple illustration on an IL6R receptor subnetwork in CellDesigner (A) and in SBGN within the Newt editor (B) and within the VANTED software (C). Source: the ACSN Dendritic Cell map: <https://acsn.curie.fr/>

Conclusion

The cd2sbgnml tool is a fully-functional bi-directional converter between the CellDesigner-specific SBML extension and SBGN-ML formats. The use of the cd2sbgnml converter allows exploration and analysis of existing collections of maps developed in CellDesigner by a large set of well-established computational Systems Biology tools specialised into the SBGN standard management such as VANTED and Newt.

Integration of the cd2sbgnml functionality into the comprehensive System Biology Format Converter framework (Rodriguez *et al.*, 2016), which manages translation between different Systems Biology standard formats including SBML, BioPAX, is planned.

Acknowledgements

We thank Marek Ostaszewski and Piotr Gawron for improving the CellDesigner version of ACSN.

Funding

This work was supported by the European Union Horizon 2020 research and innovation programme PRECISE [668858 to L.R. and E.B.], iPC [826121 to A.Z., E.B.]; and the Scientific and Technological Research Council of Turkey [111E036 to U.D.].

Conflict of Interest

None declared.

Supplementary files

Supplementary file 1 - Process ports and dissociations/associations

Illustration shown in A. CellDesigner (Funahashi *et al.*, 2007), B. Newt (<http://web.newteditor.org>) and C. VANTED (Rohn *et al.*, 2012), respectively

Supplementary file 2 - Table 1: Details on map composition (number of species and reactions involved) and conversion time for the RECON Human Metabolic network (Noronha *et al.*, 2019) and for the Atlas of Cancer Signalling Network (ACSN) (Kuperstein *et al.*, 2015).

References

- Demir,E. *et al.* (2010) The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.*, **28**, 935–942.
- Funahashi,A. *et al.* (2007) Integration of CellDesigner and SABIO-RK. *In Silico Biol.*, **7**, S81-90.
- Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinforma. Oxf. Engl.*, **19**, 524–531.
- van Iersel,M.P. *et al.* (2012) Software support for SBGN maps: SBGN-ML and LibSBGN. *Bioinforma. Oxf. Engl.*, **28**, 2016–2021.
- Kuperstein,I. *et al.* (2015) Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis*, **4**, e160.
- Le Novère,N. *et al.* (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, **34**, D689-691.
- Le Novère,N. *et al.* (2009) The Systems Biology Graphical Notation. *Nat. Biotechnol.*, **27**, 735–741.
- Mi,H. and Thomas,P. (2009) PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol. Biol. Clifton NJ*, **563**, 123–140.

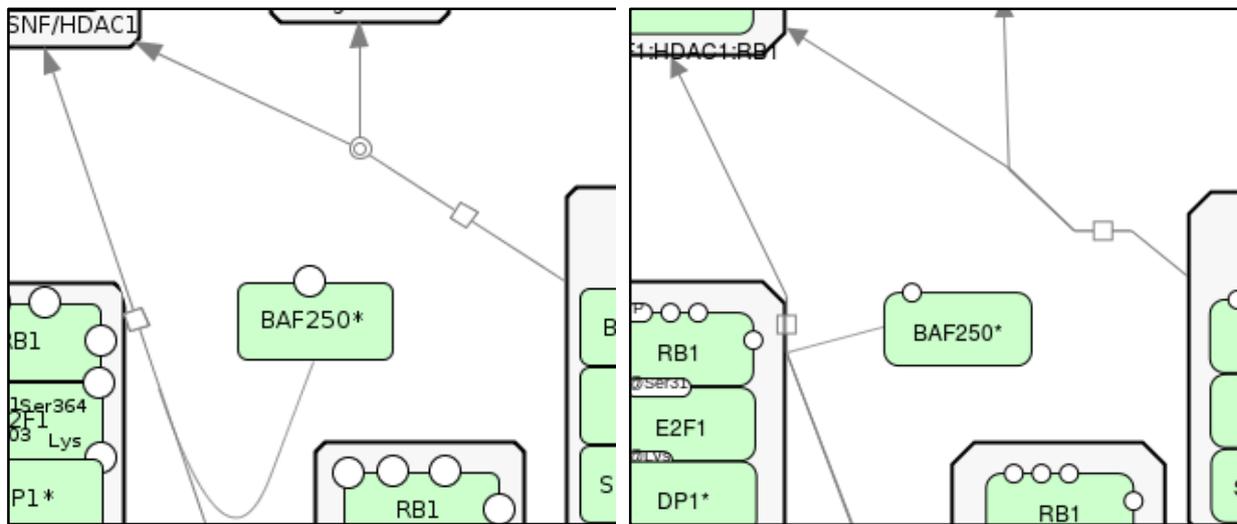
Noronha,A. *et al.* (2019) The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.*, **47**, D614–D624.

Rodriguez,N. *et al.* (2016) The systems biology format converter. *BMC Bioinformatics*, **17**, 154.

Rohn,H. *et al.* (2012) VANTED v2: a framework for systems biology applications. *BMC Syst. Biol.*, **6**, 139.

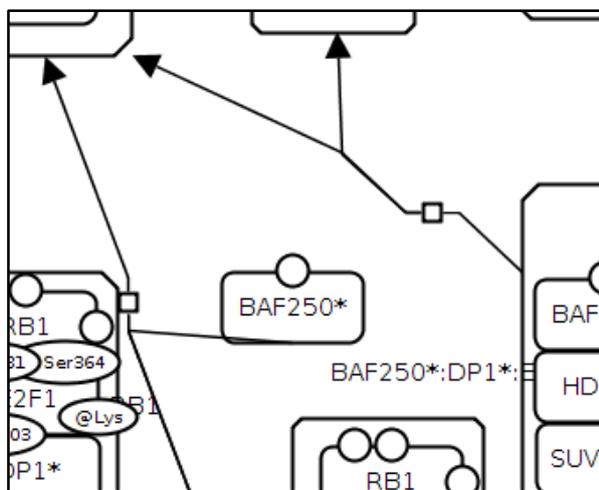
Supplementary file 1 - Process ports and dissociations/associations

Illustration shown in A. CellDesigner (Funahashi *et al.*, 2007), B. Newt (<http://web.newteditor.org/>) and C. VANTED (Rohn *et al.*, 2012) softwares, respectively



A. CellDesigner

B. Newt



C. VANTED

Supplementary file 2 - Table 1: Details on map composition (number of species and reactions involved) and conversion time for the RECON Human Metabolic network (Noronha et al., 2019) and for the Atlas of Cancer Signalling Network (ACSN) (Kuperstein et al., 2015).

The Recon3D CellDesigner file was downloaded from MINERVA interface at the Virtual Metabolic Human website, <https://www.vmh.life/#reconmap>, tab "Info", link "source file" [Accessed 29/01/2019]. Source for the ReconMap file: <https://www.vmh.life> [Accessed 29/01/2019]; source for the ACSN map collection: <https://acsncurie.fr> [Accessed 29/01/2019]. Conversion includes a) CellDesigner (CD) to SBGN and b) SBGN to CD formats. The SBGN files, generated by the cd2sbgnml converter, can be accessed at <https://acsncurie.fr/ACSN2/downloads.html> for the ACSN map collection and at <https://github.com/sbgn/cd2sbgnml/blob/master/samples/RECON.zip> for the RECON Human Metabolic network.

No. crt.	Map name	#Species	#Reactions	a) Conversion time (sec.) for CD to SBGN	b) Conversion time (sec.) for SBGN to CD
1.	Recon3D	17030	13543	13	16
2.	ReconMap 2.01	4030	5535	7	10
3.	ACSN 2.0 Master	9692	8137	18	13
4.	Cell Cycle and DNA repair map	723	544	1	1
5.	Cell Survival map	1967	1341	2	3
6.	Regulated Cell Death map	2659	2020	2	6
7.	Telomere maintenance map	153	105	1	1
8.	EMT and Senescence map	1631	1581	2	2

9.	Invasion-Motility map	390	364	1	1
10.	Angiogenesis map	237	179	1	1
11.	Adaptive Immunity map	492	338	1	1
12.	Innate Immunity map	1466	1084	1	1
13.	Cancer-Associated Fibroblasts map	682	581	1	1
14.	Dendritic Cell map	491	346	1	1
15.	Natural Killer Cell map	567	377	1	1
16.	Macrophages-MDSC map	588	457	1	1